Measurements, uncertainties and probabilistic inference/forecasting

Giulio D'Agostini

Università di Roma La Sapienza e INFN
Roma, Italy

https://www.roma1.infn.it/~dagos/AQ2021/
(temporary web page)

Falsificationism and p-values

- arXiv:physics/0412148 [physics.data-an]
- arXiv:1112.3620 [physics.data-an]
- arXiv:1609.01668 [physics.data-an]

Testing one hypothesis

- Basic Idea:
  - let's start from a 'conventional' model
    [Standard Modell, rather 'established theory', etc:]
    → "H_0" ("null hypothesis")

Testing one hypothesis

- Basic Idea:
  - let's start from a 'conventional' model
    [Standard Modell, rather 'established theory', etc:]
    → "H_0" ("null hypothesis")

  ⇒ search for violations of H_0

Testing one hypothesis

- Basic Idea:
  - let's start from a 'conventional' model
    [Standard Modell, rather 'established theory', etc:]
    → "H_0" ("null hypothesis")

  ⇒ search for violations of H_0
- Ideally

  → 'falsify'

Testing one hypothesis

- Basic Idea:
  - let's start from a 'conventional' model
    [Standard Modell, rather 'established theory', etc:]
    → "H_0" ("null hypothesis")
  - ⇒ search for violations of H_0
- Ideally
  → 'falsify'
- In practice:
  → does it make sense?
  → how is it done?

Testing <u>one</u> hypothesis

- Basic Idea:
  - let's start from a 'conventional' model
    [Standard Modell, rather 'established theory', etc:]
    → "H_0" ("null hypothesis")

  ⇒ search for violations of H_0
- Ideally

  → 'falsify'
- In practice:

  → does it make sense?

  → how is it done?

Let's review the practice and what is behind it ⇒

Falsificationism

Usually referred to Popper

and still considered by many as the *key of scientific progress*.

Falsificationism

Usually referred to Popper

and still considered by many as the *key of scientific progress*.

$$\textbf{if } C_i \not\longrightarrow E_0, \quad \textbf{then } E_0^{\textbf{(mis)}} \not\longrightarrow C_i$$

$\Rightarrow$ Causes that cannot produce the observed effects are ruled out ('falsified').

Falsificationism

Usually referred to Popper

and still considered by many as the *key of scientific progress*.

$$\textbf{if } C_i \not\rightarrow E_0, \quad \textbf{then } E_0^{(\textbf{mis})} \not\rightarrow C_i$$

⇒ Causes that cannot produce the observed effects are ruled out ('falsified').

It seems OK – *'obvious'*! – but it is indeed naïve for several aspects.

Proof by contradiction . . . ‘extended’. . .

Falsification rule: to what is it ‘inspired’?

Proof by contradiction . . . 'extended'. . .

Falsification rule: to what is it 'inspired'?

Proof by contradiction of classical, deductive logic:

- Assume that a hypothesis is true;
- Derive 'all' logical consequences;
- If (at least) one of the consequences is known to be false, then the hypothesis is rejected.

Proof by contradiction . . . 'extended'. . .

Falsification rule: to what is it 'inspired'?

Proof by contradiction of classical, deductive logic:

- Assume that a hypothesis is true;
- Derive 'all' logical consequences;
- If (at least) one of the consequences is known to be false, then the hypothesis is rejected.

Popperian falsificationism

extends the reasoning to experimental sciences

Proof by contradiction . . . ‘extended’. . .

Falsification rule: to what is it ‘inspired’?

Proof by contradiction of classical, deductive logic:

- Assume that a hypothesis is true;
- Derive ‘all’ logical consequences;
- If (at least) one of the consequences is known to be false, then the hypothesis is rejected.

Popperian falsificationism

extends the reasoning to experimental sciences

is this extension legitimate?

Falsificationism? OK, but...

- What shall we do of all hypotheses not yet falsified? (Limbus? How should we progress?)

Falsificationism? OK, but. . .

- What shall we do of all hypotheses not yet falsified? (Limbus? How should we progress?)
- What to do if nothing of what can be observed is incompatible with the hypothesis (or with many hypotheses)?

Falsificationism? OK, but...

- What shall we do of all hypotheses not yet falsified? (Limbus? How should we progress?)
- What to do if nothing of what can be observed is incompatible with the hypothesis (or with many hypotheses)?

  E.g. H_i being a Gaussian $f(x \,|\, \mu_i, \sigma_i)$

  $\Rightarrow$ Given any pair or parameters $\{\mu_i, \sigma_i\}$ (i.e. $\forall H_i$), all values of x from $-\infty$ to $+\infty$ are possible.

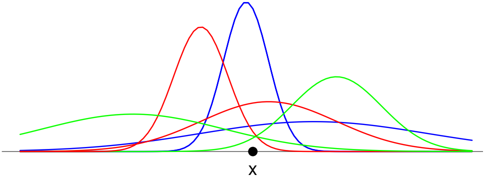
Falsificationism? OK, but. . .

- What shall we do of all hypotheses not yet falsified? (Limbus? How should we progress?)
- What to do if nothing of what can be observed is incompatible with the hypothesis (or with many hypotheses)?

  E.g. H_i being a Gaussian $f(x \,|\, \mu_i, \sigma_i)$

  $\Rightarrow$ Given any pair or parameters $\{\mu_i, \sigma_i\}$ (i.e. $\forall H_i$), all values of x from $-\infty$ to $+\infty$ are possible.

  $\Rightarrow$ Having observed any value **x**, none of H_i can be, strictly speaking, falsified.



Falsificationism in action. . .

Obviously, this does not means that falsificationism is never applicable,

Falsificationism in action. . .

Obviously, this does not means that falsificationism is never applicable, but as long as no stochastic processes are involved (randomness inherent to the physical processes, or due to ‘errors’ in measurement).

Falsificationism in action. . .

Obviously, this does not means that falsificationism is never applicable, but as long as no stochastic processes are involved (randomness inherent to the physical processes,
or due to ‘errors’ in measurement).

$\Rightarrow$ Practically never in the experimental sciences!

Falsificationism in action...

Obviously, this does not means that falsificationism is never applicable, but as long as no stochastic processes are involved (randomness inherent to the physical processes, or due to 'errors' in measurement).

⇒ Practically never in the experimental sciences!

- Science proceeds, in practice, rather differently:

  The natural development of Science shows that researches are carried along the directions that seem more credible (and hopefully fruitful) at a given moment.

Falsificationism in action...

Obviously, this does not means that falsificationism is never applicable, but as long as no stochastic processes are involved (randomness inherent to the physical processes, or due to 'errors' in measurement).

⇒ Practically never in the experimental sciences!

- Science proceeds, in practice, rather differently:

  The natural development of Science shows that researches are carried along the directions that seem more credible (and hopefully fruitful) at a given moment. A behavior "179 degrees or so out of phase from Popper's idea that we make progress by falsificating theories"
  (Wilczek, `http://arxiv.org/abs/physics/0403115` *)*

Falsificationism in action...

Obviously, this does not means that falsificationism is never applicable, but as long as no stochastic processes are involved (randomness inherent to the physical processes,
or due to 'errors' in measurement).

⇒ Practically never in the experimental sciences!

- Science proceeds, in practice, rather differently:

  The natural development of Science shows that researches are carried along the directions that seem more <u>credible</u> (and hopefully fruitful) at a given moment. A behavior "179 degrees or so out of phase from Popper's idea that we make progress by falsificating theories" (*Wilczek*, `http://arxiv.org/abs/physics/0403115`)

⇒ logically speaking, Popper's falsificationism
has to be considered ... falsified!

Falsificationism and statistics

...then, statisticians have invented the "hypothesis tests"

Falsificationism and statistics

. . . then, statisticians have invented the "hypothesis tests",
in which the impossible is replaced by the improbable!

Falsificationism and statistics

...then, statisticians have invented the "hypothesis tests",
in which the impossible is replaced by the improbable!

But from the 'impossible' to the 'improbable' there is not just a question of quantity, but a question of quality.

Falsificationism and statistics

…then, statisticians have invented the "hypothesis tests", in which the impossible is replaced by the improbable!

But from the 'impossible' to the 'improbable' there is not just a question of quantity, but a question of quality.

This mechanism, logically flawed, is particularly dangerous because is deeply rooted in most scientists, due to education and custom, although not supported by logic.

Falsificationism and statistics

…then, statisticians have invented the "hypothesis tests",
in which the impossible is replaced by the improbable!

But from the 'impossible' to the 'improbable' there is not just a question of quantity, but a question of quality.

This mechanism, logically flawed, is particularly dangerous because is deeply rooted in most scientists, due to education and custom, although not supported by logic.

⇒ Basically responsible of all fake claims of discoveries in the past decades.

Falsificationism and statistics

...then, statisticians have invented the "hypothesis tests",
in which the impossible is replaced by the improbable!

But from the 'impossible' to the 'improbable' there is not just a question of quantity, but a question of quality.

This mechanism, logically flawed, is particularly dangerous because is deeply rooted in most scientists, due to education and custom, although not supported by logic.

⇒ Basically responsible of all fake claims of discoveries in the past decades.

[*I am particularly worried about claims concerning our health, or the status of the Planet, etc. ...*]

In summary

A) **if** $C_i \not\rightarrow E$, and **we observe** E
$\Rightarrow C_i$ is impossible ('false')

In summary

A) **if** $C_i \not\rightarrow E$, and **we observe** E
$\Rightarrow C_i$ is impossible ('false')

B) **if** $C_i \xrightarrow[\text{small probability}]{} E$, and **we observe** E

$\Rightarrow C_i$ has small probability to be true
"most likely false"

In summary

A) **if** $C_i \not\longrightarrow E$, and **we observe** E OK

$\Rightarrow C_i$ is impossible ('false')

B) **if** $C_i \xrightarrow[\text{small probability}]{} E$, and **we observe** E

$\Rightarrow C_i$ has small probability to be true

"most likely false"

In summary

A) **if** $C_i \not\longrightarrow E$, and **we observe** E OK

$\Rightarrow C_i$ is impossible ('false')

B) ~~**if** $C_i \xrightarrow[\text{small probability}]{} E$, and **we observe** E~~ NO

~~$\Rightarrow C_i$ has small probability to be true~~

~~"most likely false"~~

But it is behind the rational behind the statistical hypothesis tests!

$P(A \,|\, B) \leftrightarrow P(B \,|\, A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \,|\, B) \neq P(B \,|\, A)$

$P(A \,|\, B) \leftrightarrow P(B \,|\, A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \,|\, B) \neq P(B \,|\, A)$

- $P(\text{Positive} \,|\, \overline{HIV}) \neq P(\overline{HIV} \,|\, \text{Positive})$

$P(A \,|\, B) \leftrightarrow P(B \,|\, A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \,|\, B) \neq P(B \,|\, A)$

- $P(\text{Positive} \,|\, \overline{HIV}) \neq P(\overline{HIV} \,|\, \text{Positive})$
- $P(\text{Win} \,|\, \text{Play}) \neq P(\text{Play} \,|\, \text{Win})$ [Lotto]

$P(A \mid B) \leftrightarrow P(B \mid A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \mid B) \neq P(B \mid A)$

- $P(\text{Positive} \mid \overline{HIV}) \neq P(\overline{HIV} \mid \text{Positive})$
- $P(\text{Win} \mid \text{Play}) \neq P(\text{Play} \mid \text{Win})$ [Lotto]
- $P(\text{Pregnant} \mid \text{Woman}) \neq P(\text{Woman} \mid \text{Pregnant})$

$P(A \,|\, B) \leftrightarrow P(B \,|\, A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \,|\, B) \neq P(B \,|\, A)$

- $P(\text{Positive} \,|\, \overline{HIV}) \neq P(\overline{HIV} \,|\, \text{Positive})$
- $P(\text{Win} \,|\, \text{Play}) \neq P(\text{Play} \,|\, \text{Win})$ [Lotto]
- $P(\text{Pregnant} \,|\, \text{Woman}) \neq P(\text{Woman} \,|\, \text{Pregnant})$

In particular

- A cause might produce a given effect with very low probability, and nevertheless could be the most probable cause of that effect

$P(A \mid B) \leftrightarrow P(B \mid A)$

Pay attention not to arbitrary revert conditional probabilities:

In general $P(A \mid B) \neq P(B \mid A)$

- $P(\text{Positive} \mid \overline{HIV}) \neq P(\overline{HIV} \mid \text{Positive})$
- $P(\text{Win} \mid \text{Play}) \neq P(\text{Play} \mid \text{Win})$ [Lotto]
- $P(\text{Pregnant} \mid \text{Woman}) \neq P(\text{Woman} \mid \text{Pregnant})$

In particular

- A cause might produce a given effect with very low probability, and nevertheless could be the most probable cause of that effect, **often the only one!**

‘Low probability’ events

Typical values of statistical practice to ‘reject’ a hypothesis are 5%, 1%, . . .

‘Low probability’ events

Typical values of statistical practice to ‘reject’ a hypothesis are 5%, 1%, ...

BUT the greatest majority of the events of interest have very low probability (before occurring!).

'Low probability' events

Typical values of statistical practice to 'reject' a hypothesis are 5%, 1%, ...

BUT the greatest majority of the events of interest have very low probability (before occurring!).

For example, imagine a Gaussian random generator (H_0, with $\mu = 3, \sigma = 1$) gives us $x_{obs} = 3.1416$.

'Low probability' events

Typical values of statistical practice to 'reject' a hypothesis are 5%, 1%, ...

BUT the greatest majority of the events of interest have very low probability (before occurring!).

For example, imagine a Gaussian random generator (H_0, with $\mu = 3, \sigma = 1$) gives us $x_{obs} = 3.1416$.

→ What 'was' the probability to give exactly that number?

'Low probability' events

Typical values of statistical practice to 'reject' a hypothesis are 5%, 1%, ...

BUT the greatest majority of the events of interest have very low probability (before occurring!).

For example, imagine a Gaussian random generator (H_0, with $\mu = 3, \sigma = 1$) gives us $x_{obs} = 3.1416$.

$\rightarrow$ What 'was' the probability to give exactly that number?

$$
\begin{aligned}
P(x_{obs} = 3.1416 \,|\, H_0) &= \int_{3.14155}^{3.14165} f_{\mathcal{G}}(x \,|\, \mu, \sigma)dx \\
&\approx f_{\mathcal{G}}(3.1416 \,|\, \mu, \sigma) \times 0.0001 \\
&\approx 39 \times 10^{-6}
\end{aligned}
$$

'Low probability' events

Typical values of statistical practice to 'reject' a hypothesis are 5%, 1%, ...

BUT the greatest majority of the events of interest have very low probability (before occurring!).

For example, imagine a Gaussian random generator (H_0, with $\mu = 3, \sigma = 1$) gives us $x_{obs} = 3.1416$.

→ What 'was' the probability to give exactly that number?

$$
\begin{aligned}
P(x_{obs} = 3.1416 \,|\, H_0) &= \int_{3.14155}^{3.14165} f_{\mathcal{G}}(x \,|\, \mu, \sigma)\, dx \\
&\approx f_{\mathcal{G}}(3.1416 \,|\, \mu, \sigma) \times 0.0001 \\
&\approx 39 \times 10^{-6}
\end{aligned}
$$

→ What is the probability that x_{obs} comes from H_0?

'Low probability' events

Typical values of statistical practice to 'reject' a hypothesis are 5%, 1%, . . .

BUT the greatest majority of the events of interest have very low probability (before occurring!).

For example, imagine a Gaussian random generator (H_0, with $\mu = 3, \sigma = 1$) gives us $x_{obs} = 3.1416$.

→ What 'was' the probability to give exactly that number?

$$
\begin{aligned}
P(x_{obs} = 3.1416 \,|\, H_0) &= \int_{3.14155}^{3.14165} f_{\mathcal{G}}(x \,|\, \mu, \sigma)\, dx \\
&\approx f_{\mathcal{G}}(3.1416 \,|\, \mu, \sigma) \times 0.0001 \\
&\approx 39 \times 10^{-6}
\end{aligned}
$$

→ What is the probability that x_{obs} comes from H_0?

- Certainly NOT $\approx 39 \times 10^{-6}$;

Most events 'had' very small probability to occur!

$$
\begin{aligned}
P(x_{obs} = 3.1416 \,|\, H_0) &\approx 39 \times 10^{-6} \\
P(H_0 \,|\, x_{obs} = 3.1416) &= 1\,.
\end{aligned}
$$

Most events ‘had’ very small probability to occur!

$$\begin{aligned} P(x_{obs} = 3.1416 \,|\, H_0) &\approx 39 \times 10^{-6} \\ P(H_0 \,|\, x_{obs} = 3.1416) &= 1\,. \end{aligned}$$

Other, real life example:

- I shut a picture with my faithful pocket camera.

Most events 'had' very small probability to occur!

$$\begin{aligned} P(x_{obs} = 3.1416 \,|\, H_0) &\approx 39 \times 10^{-6} \\ P(H_0 \,|\, x_{obs} = 3.1416) &= 1\,. \end{aligned}$$

Other, real life example:

- I shut a picture with my faithful pocket camera.
- What is the probability of every configuration of the three RGB codes of the 20MB pixels, given this scene?

$$P(\text{Picture} \equiv \text{X}_{recorded} \,|\, \text{This scene}) \lll 1$$

Most events 'had' very small probability to occur!

$$
\begin{aligned}
P(x_{obs} = 3.1416 \,|\, H_0) &\approx 39 \times 10^{-6} \\
P(H_0 \,|\, x_{obs} = 3.1416) &= 1\,.
\end{aligned}
$$

Other, real life example:

- I shut a picture with my faithful pocket camera.
- What is the probability of every configuration of the three RGB codes of the 20MB pixels, given this scene?

$$P(\text{Picture} \equiv \text{X}_{recorded} \,|\, \text{This scene}) \lll 1$$

- But

$$P(\text{This scene} \,|\, \text{Picture}) = 1$$

Most events 'had' very small probability to occur!

$$\begin{aligned} P(x_{obs} = 3.1416 \,|\, H_0) &\approx 39 \times 10^{-6} \\ P(H_0 \,|\, x_{obs} = 3.1416) &= 1\,. \end{aligned}$$

Other, real life example:

- I shut a picture with my faithful pocket camera.
- What is the probability of every configuration of the three RGB codes of the 20MB pixels, given this scene?

$$P(\text{Picture} \equiv \text{X}_{recorded} \,|\, \text{This scene}) \lll 1$$

- But

$$P(\text{This scene} \,|\, \text{Picture}) = 1$$

What else?

An so on...

Probability of something else. . .

Besides the logical flow, the 'technical issue' of low probability events which would lead to reject any hypothesis forces the statisticians to rethink the question. . .

Probability of something else...

Besides the logical flow, the 'technical issue' of low probability events which would lead to reject any hypothesis forces the statisticians to rethink the question...

> but, instead of repent, throw everything away and finally start to read Laplace, they made a new invention:

Probability of something else...

Besides the logical flow, the 'technical issue' of low probability events which would lead to reject any hypothesis forces the statisticians to rethink the question...

> but, instead of repent, throw everything away and finally start to read Laplace, they made a new invention:
> → what matters is not the probability of the x_{obs}, but rather the probability of x_{obs} or of any other *less probable* value:

Probability of something else...

Besides the logical flow, the 'technical issue' of low probability events which would lead to reject any hypothesis forces the statisticians to rethink the question...

> but, instead of repent, throw everything away and finally start to read Laplace, they made a new invention:
> → what matters is not the probability of the x_{obs}, but rather the probability of x_{obs} or of any other *less probable* value:
>
> $$P(X \geq 3.1416) \;=\; \int_{3.14155}^{+\infty} f_{\mathcal{G}}(x \,|\, \mu, \sigma)\, dx \approx 44\%$$

Probability of something else...

Besides the logical flow, the 'technical issue' of low probability events which would lead to reject any hypothesis forces the statisticians to rethink the question...

> but, instead of repent, throw everything away and finally start to read Laplace, they made a new invention:
> → what matters is not the probability of the x_{obs}, but rather the probability of x_{obs} or of any other *less probable* value:
>
> $$P(X \geq 3.1416) = \int_{3.14155}^{+\infty} f_{\mathcal{G}}(x \,|\, \mu, \sigma)\, dx \approx 44\%$$
>
> $$P(X \geq x_{obs}) \Rightarrow \text{'p-value'}$$

Probability of something else...

Besides the logical flow, the 'technical issue' of low probability events which would lead to reject any hypothesis forces the statisticians to rethink the question...

but, instead of repent, throw everything away and finally start to read Laplace, they made a new invention:
→ what matters is not the probability of the x_{obs}, but rather the probability of x_{obs} or of any other *less probable* value:

$$
\begin{aligned}
P(X \geq 3.1416) &= \int_{3.14155}^{+\infty} f_{\mathcal{G}}(x \,|\, \mu, \sigma)\, dx \approx 44\% \\
P(X \geq x_{obs}) &\Rightarrow \text{`p-value'}
\end{aligned}
$$

⇒ Magically the 'result' becomes rather probable!
Why, we, silly, worried about it?

Probability of something else...

Besides the logical flow, the 'technical issue' of low probability events which would lead to reject any hypothesis forces the statisticians to rethink the question...

but, instead of repent, throw everything away and finally start to read Laplace, they made a new invention:
$\rightarrow$ what matters is not the probability of the x_{obs}, but rather the probability of x_{obs} or of any other *less probable* value:

$$P(X \geq 3.1416) = \int_{3.14155}^{+\infty} f_{\mathcal{G}}(x \,|\, \mu, \sigma)\,dx \approx 44\%$$

$$P(X \geq x_{obs}) \Rightarrow \text{`p-value'}$$

$\Rightarrow$ Magically the 'result' becomes rather probable!
Why, we, silly, worried about it?

$\Rightarrow$ 'Statisticians' are happy...

Probability of something else...

Besides the logical flow, the 'technical issue' of low probability events which would lead to reject any hypothesis forces the statisticians to rethink the question...

but, instead of repent, throw everything away and finally start to read Laplace, they made a new invention:
→ what matters is not the probability of the x_{obs}, but rather the probability of x_{obs} or of any other *less probable* value:

$$
\begin{aligned}
P(X \geq 3.1416) &= \int_{3.14155}^{+\infty} f_{\mathcal{G}}(x \,|\, \mu, \sigma)\, dx \approx 44\% \\
P(X \geq x_{obs}) &\Rightarrow \text{'p-value'}
\end{aligned}
$$

⇒ Magically the 'result' becomes rather probable!
Why, we, silly, worried about it?

⇒ 'Statisticians' are happy...
scientists and general public get cheated...

Probability of something else...

Besides the logical flow, the 'technical issue' of low probability events which would lead to reject any hypothesis forces the statisticians to rethink the question...

but, instead of repent, throw everything away and finally start to read Laplace, they made a new invention:
$\rightarrow$ what matters is not the probability of the x_{obs}, but rather the probability of x_{obs} or of any other *less probable* value:

$$P(X \geq 3.1416) = \int_{3.14155}^{+\infty} f_{\mathcal{G}}(x \,|\, \mu, \sigma)\, dx \approx 44\%$$

$$P(X \geq x_{obs}) \Rightarrow \text{'p-value'}$$

$\Rightarrow$ Magically the 'result' becomes rather probable!
Why, we, silly, worried about it?

$\Rightarrow$ 'Statisticians' are happy...
scientists and general public get cheated...
(From the logical point of view the situation gets worsened:
$\rightarrow$ conclusions depend on events not actually observed!)

Which p-value?...

'p-value' = 'probability of the tail(s)'

Which p-value?...

'p-value' = 'probability of the tail(s)'

Of what?

Which p-value?...

'p-value' = 'probability of the tail(s)'

Of what?

→ the test variable ('θ') is absolutely arbitrary:

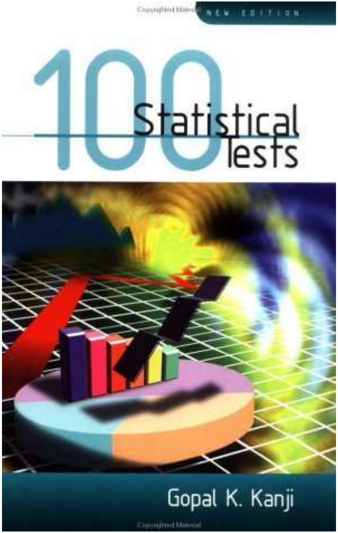
$$\theta = \theta(\mathrm{x})$$

$$\rightarrow f(\theta) \text{ [p.d.f]}$$

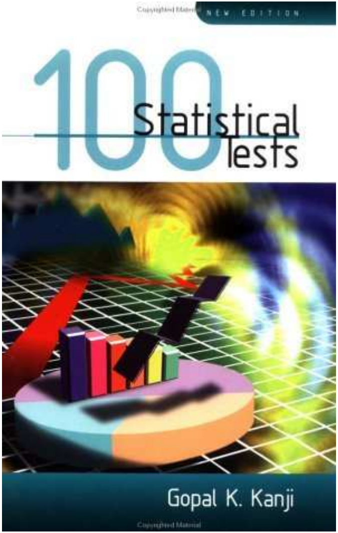
$$\text{Experiment:} \rightarrow \theta_{obs} = \theta(\mathrm{x}_{obs})$$

$$\text{p-value} = P(\theta \geq \theta_{obs}) \qquad \text{('one tail')}$$

Which p-value?. . .

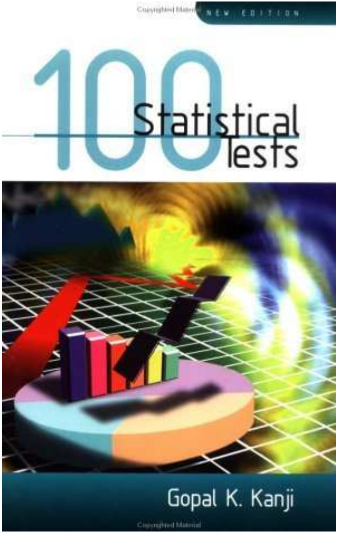


Which p-value?...



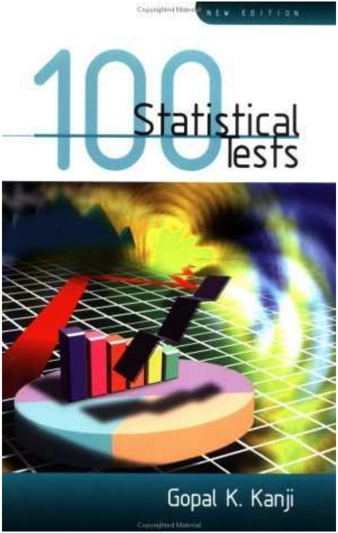
- far from exhaustive list,

Which p-value?...



- far from exhaustive list,
- with arbitrary variants:

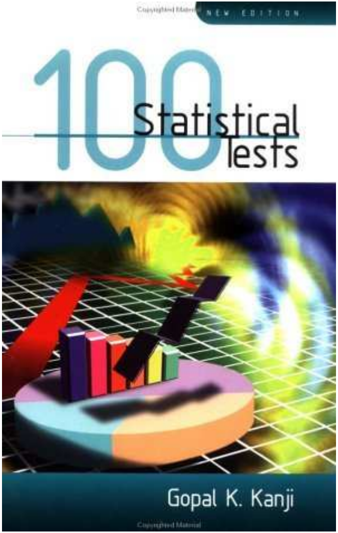
Which p-value?...



- far from exhaustive list,
- with arbitrary variants:

⇒ practitioners chose the one that provide the result they like better:
→ *like if you go around until "someone agrees with you"*

Which p-value?...



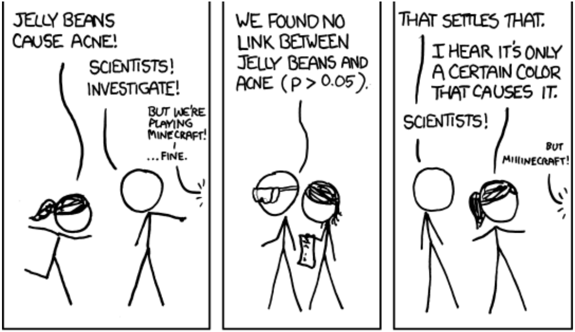
- far from exhaustive list,
- with arbitrary variants:

⇒ practitioners chose the one that provide the result they like better:
→ *like if you go around until "someone agrees with you"*

- personal **'golden rule'**:
"the more exotic is the name of the test, the less I believe the result", because I'm pretty sure that several 'normal' tests have been discarded in the meanwhile...

Or look around, searching for ‘significance’

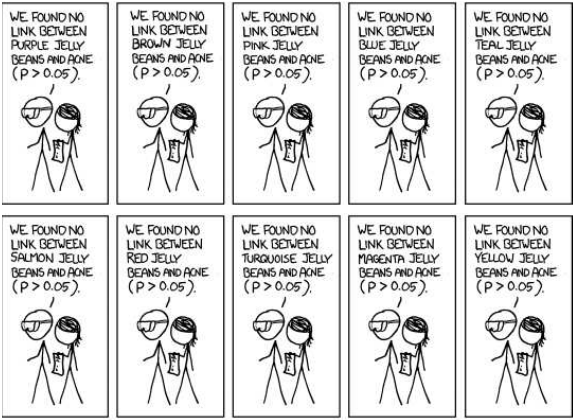
If changing the test does not help, change hypotheses...



[http://xkcd.com/882/]

Or look around, searching for 'significance'

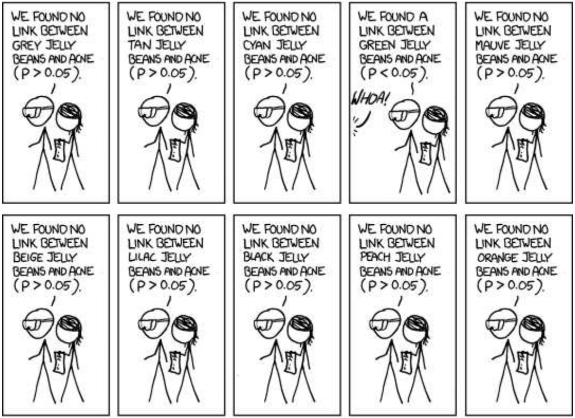
If changing the test does not help, change hypotheses...



[http://xkcd.com/882/]

Or look around, searching for 'significance'

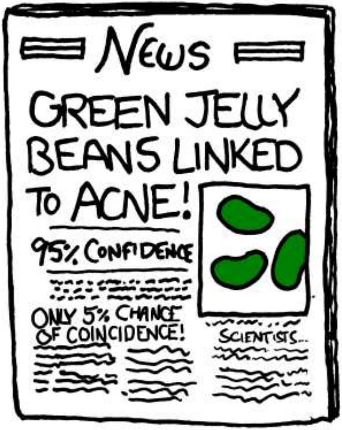
If changing the test does not help, change hypotheses...



[http://xkcd.com/882/]

Or look around, searching for‘significance’

If changing the test does not help, change hypotheses...



[http://xkcd.com/882/]

P-hacking ("p-value hacking")

The 'science' of inventing significant results. . .

p-hacking, or cheating on a p-value

June 11, 2015

By arthur charpentier

Share

(This article was first published on **Freakonometrics » R-english**, and kindly contributed to R-bloggers)

Yesterday evening, I discovered some interesting slides on False-Positives, p-Hacking, Statistical Power, and Evidential Value, via @UCBITSS 's post on Twitter. More precisely, there was this slide on how cheating (because that's basically what it is) to get a 'good' model (by targeting the *p*-value)

1. Stop collecting data once $p<.05$
2. Analyze many measures, but report only those with $p<.05$.
3. Collect and analyze many conditions, but only report those with $p<.05$.
4. Use covariates to get $p<.05$.
5. Exclude participants to get $p<.05$.
6. Transform the data to get $p<.05$.

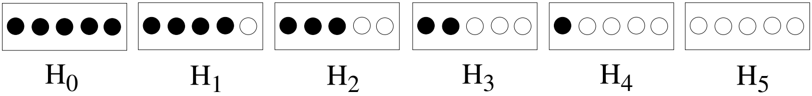
http://www.r-bloggers.com/p-hacking-or-cheating-on-a-p-value/

- Google for "p-hacking"

Continuing

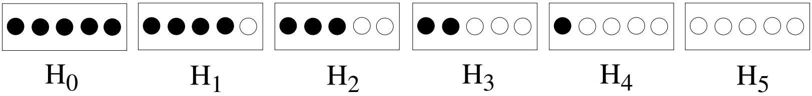
from last lecture

Which box? Which ball?



Let us take randomly one of the boxes.

Which box? Which ball?



Let us take randomly one of the boxes.

We are in a state of uncertainty concerning several *events*, the most important of which correspond to the following questions:

(a) Which box have we chosen, H_0, H_1, ..., H_5?

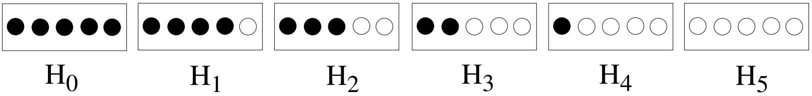
(b) If we extract randomly a ball from the chosen box, will we observe a white ($E_W \equiv E_1$) or black ($E_B \equiv E_2$) ball?

Our certainties:

$$\cup_{j=0}^{5} H_j = \Omega$$

$$\cup_{i=1}^{2} E_i = \Omega\,.$$

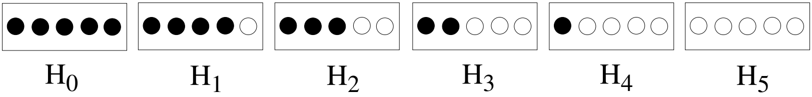
Which box? Which ball?



Let us take randomly one of the boxes.

- What happens after we have extracted one ball and looked its color?
  - Intuitively feel *how to* roughly *change* our opinion about
    - the possible cause
    - a future observation

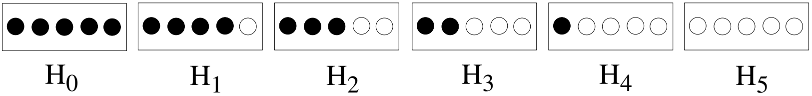
Which box? Which ball?



Let us take randomly one of the boxes.

- What happens after we have extracted one ball and looked its color?
  - Intuitively feel *how to* roughly *change* our opinion about
    - the possible cause
    - a future observation
  - Can we do it *quantitatively*, in an ‘objective way’?

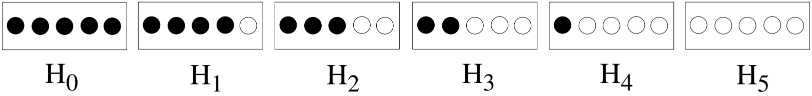
Which box? Which ball?



Let us take randomly one of the boxes.

- What happens after we have extracted one ball and looked its color?
  - Intuitively feel *how to* roughly *change* our opinion about
    - the possible cause
    - a future observation
  - Can we do it *quantitatively*, in an 'objective way'?
- And after a sequence of extractions?

Which box? Which ball?



Let us take randomly one of the boxes.

- What happens after we have extracted one ball and looked its color?
  - Intuitively feel *how to* roughly *change* our opinion about
    - the possible cause
    - a future observation
  - Can we do it *quantitatively*, in an 'objective way'?
- And after a sequence of extractions?

Note: In general, we are uncertain about all the combinations of E_i and H_j:

$$E_1 \cap H_0,\ E_1 \cap H_1,\ \ldots,\ E_2 \cap H_5,$$

and these 12 constituents are not equiprobable.

Subjective nature of probability

"Since the knowledge may be different with different persons

Subjective nature of probability

"Since the knowledge may be different with different persons or with the same person at different times,

Subjective nature of probability

"Since the knowledge may be different with different persons or with the same person at different times, they may anticipate the same event with more or less confidence,

Subjective nature of probability

"Since the knowledge may be different with different persons or with the same person at different times, they may anticipate the same event with more or less confidence, and thus different numerical probabilities may be attached to the same event"

Subjective nature of probability

"Since the knowledge may be different with different persons or with the same person at different times, they may anticipate the same event with more or less confidence, and thus different numerical probabilities may be attached to the same event"

(Schrödinger, 1947)

Subjective nature of probability

"Since the knowledge may be different with different persons or with the same person at different times, they may anticipate the same event with more or less confidence, and thus different numerical probabilities may be attached to the same event"

(Schrödinger, 1947)

Probability depends on the status of information of the *subject* who evaluates it.

Probability is always conditional probability

"Thus whenever we speak loosely of 'the probability of an event', it is always to be understood: probability with regard to a certain given state of knowledge"

Probability is always conditional probability

"Thus whenever we speak loosely of 'the probability of an event', it is always to be understood: probability with regard to a certain given state of knowledge"

(Schrödinger, 1947)

Probability is always conditional probability

"Thus whenever we speak loosely of 'the probability of an event', it is always to be understood: probability with regard to a certain given state of knowledge"

(Schrödinger, 1947)

$$P(E) \longrightarrow P(E \mid I_s(t))$$

where $I_s(t)$ is the information available to *subject* s at time t.

Probability is always conditional probability

"Thus whenever we speak loosely of 'the probability of an event', it is always to be understood: probability with regard to a certain given state of knowledge"

(Schrödinger, 1947)

$$P(E) \quad \longrightarrow \quad P(E \,|\, I_s(t))$$

where $I_s(t)$ is the information available to *subject* s at time t.

Examples:

- tossing coins and dice;
- the three box problem.

What are we talking about?

"Given the state of our knowledge about everything that could possible have any bearing on the coming true...

What are we talking about?

"Given the state of our knowledge about everything that could possible have any bearing on the coming true... the numerical probability P of this event is to be a real number by the indication of which we try in some cases to setup a quantitative measure of the strength of our conjecture or anticipation, founded on the said knowledge, that the event comes true"

(Schrödinger, 1947)

What are we talking about?

"Given the state of our knowledge about everything that could possible have any bearing on the coming true. . . the numerical probability P of this event is to be a real number by the indication of which we try in some cases to setup a quantitative measure of the strength of our conjecture or anticipation, founded on the said knowledge, that the event comes true"

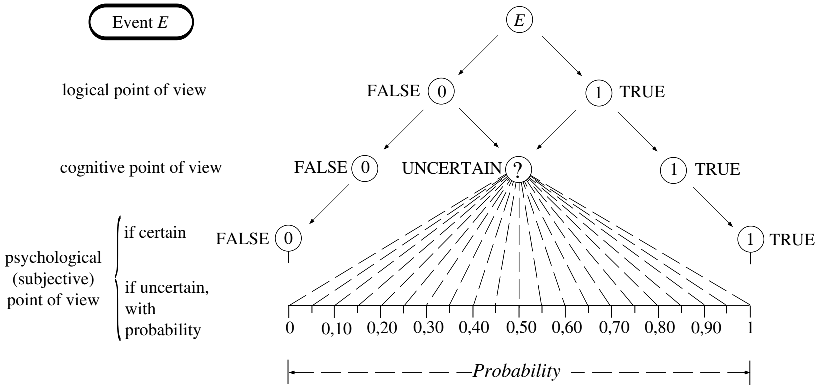
$\Rightarrow$ How much we believe something

What are we talking about?

"Given the state of our knowledge about everything that could possible have any bearing on the coming true. . . the numerical probability P of this event is to be a real number by the indication of which we try in some cases to setup a quantitative measure of the strength of our conjecture or anticipation, founded on the said knowledge, that the event comes true"

$\rightarrow$ 'Degree of belief' $\leftarrow$

False, True and probable



Beliefs and 'coherent' bets

Remarks:

- Subjective does not mean arbitrary!

Beliefs and ‘coherent’ bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?

Beliefs and 'coherent' bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet

Beliefs and ‘coherent’ bets

Remarks:

- Subjective <u>does not mean</u> arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet:
  - you state the odds according on your beliefs;
  - somebody else will choose the direction of the bet.

Beliefs and ‘coherent’ bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet:
  - you state the odds according on your beliefs;
  - somebody else will choose the direction of the bet.

“His [Bouvard] calculations give him the mass of Saturn as 3,512th part of that of the sun. Applying my probabilistic formulae to these observations, I find that the odds are 11,000 to 1 that the error in this result is not a hundredth of its value.” (Laplace)

Beliefs and 'coherent' bets

Remarks:

- Subjective does not mean arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet:
  - you state the odds according on your beliefs;
  - somebody else will choose the direction of the bet.

"His [Bouvard] calculations give him the mass of Saturn as 3,512th part of that of the sun. Applying my probabilistic formulae to these observations, I find that the odds are 11,000 to 1 that the error in this result is not a hundredth of its value." (Laplace)

$\rightarrow P(3477 \leq M_{Sun}/M_{Sat} \leq 3547 \,|\, I(\text{Laplace})) = 99.99\%$

Beliefs and ‘coherent’ bets

Remarks:

- Subjective <u>does not mean</u> arbitrary!
- How to force people to assess how much they are confident on something?
  - Coherent bet:
  - you state the odds according on your beliefs;
  - somebody else will choose the direction of the bet.

“His [Bouvard] calculations give him the mass of Saturn as 3,512th part of that of the sun. Applying my probabilistic formulae to these observations, I find that the odds are 11,000 to 1 that the error in this result is not a hundredth of its value.” (Laplace)

$\rightarrow P(3477 \leq M_{Sun}/M_{Sat} \leq 3547 \,|\, I(\text{Laplace})) = 99.99\%$

Is a ‘conventional’ 95% C.L. lower/upper bound a 19 to 1 bet?

Standard textbook definitions

$$p = \frac{\#\,\text{favorable cases}}{\#\,\text{possible equiprobable cases}}$$

$$p = \frac{\#\,\text{times the event has occurred}}{\#\,\text{independent trials under same conditions}}$$

Standard textbook definitions

It is easy to check that 'scientific' definitions suffer of circularity

$$p = \frac{\#\,\text{favorable cases}}{\#\,\text{possible equiprobable cases}}$$

$$p = \frac{\#\,\text{times the event has occurred}}{\#\,\text{independent trials under same conditions}}$$

Standard textbook definitions

It is easy to check that 'scientific' definitions suffer of circularity

$$p = \frac{\#\,\text{favorable cases}}{\#\,\text{possible equally possible cases}}$$

$$p = \frac{\#\,\text{times the event has occurred}}{\#\,\text{independent trials under same conditions}}$$

Note!: *"lorsque rien ne porte à croire que l'un de ces cas doit arriver plutot que les autres" (Laplace)*

Replacing 'equi-probable' by 'equi-possible' is just cheating students (as I did in my first lecture on the subject...).

Standard textbook definitions

It is easy to check that 'scientific' definitions suffer of circularity, plus other problems

$$p = \frac{\#\,\text{favorable cases}}{\#\,\text{possible equiprobable cases}}$$

$$p = \lim_{n\rightarrow\infty} \frac{\#\,\text{times the event has occurred}}{\#\,\text{independent trials under same condition}}$$

Future $\Leftrightarrow$ Past (belief!)

$n \rightarrow \infty$: $\rightarrow$ *"usque tandem?"*

$\rightarrow$ *"in the long run we are all dead"*

$\rightarrow$ It limits the range of applications

Standard textbook definitions

It is easy to check that 'scientific' definitions suffer of circularity, plus other problems

$$p = \frac{\#\text{ favorable cases}}{\#\text{ possible equiprobable cases}}$$

$$p = \lim_{n\rightarrow\infty} \frac{\#\text{ times the event has occurred}}{\#\text{ trials under}}$$

Future ⇔ Past (belief!)

$n \rightarrow \infty$: → *"usque tandem?"*

→ *"in the long run we are all dead"*

→ It limits the range of applications

Future ⇔ Past: avoid the end of the *inductivist turkey*!

‘Definitions’ $\rightarrow$ evaluation rules

Very useful evaluation rules

$$A) \quad p \;=\; \frac{\#\,\text{favorable cases}}{\#\,\text{possible equiprobable cases}}$$

$$B) \quad p \;=\; \frac{\#\,\text{times the event has occurred}}{\#\text{independent trials under same condition}}$$

If the implicit beliefs are well suited for each case of application.

‘Definitions’ $\rightarrow$ evaluation rules

Very useful evaluation rules

$$A) \quad p = \frac{\#\,\text{favorable cases}}{\#\,\text{possible equiprobable cases}}$$

$$B) \quad p = \frac{\#\,\text{times the event has occurred}}{\#\text{independent trials under same condition}}$$

If the implicit beliefs are well suited for each case of application.

BUT they cannot define the concept of probability!

‘Definitions’ $\rightarrow$ evaluation rules

Very useful evaluation rules

$$A) \quad p = \frac{\#\,\text{favorable cases}}{\#\,\text{possible equiprobable cases}}$$

$$B) \quad p = \frac{\#\,\text{times the event has occurred}}{\#\text{independent trials under same condition}}$$

In the probabilistic approach we are following

- Rule A is recovered immediately (under the assumption of equiprobability, when it applies).
- Rule B results from a theorem of Probability Theory (under well defined assumptions).

‘Definitions’ → evaluation rules

Very useful evaluation rules

$$A) \quad p \;=\; \frac{\#\,\text{favorable cases}}{\#\,\text{possible equiprobable cases}}$$

$$B) \quad p \;=\; \frac{\#\,\text{times the event has occurred}}{\#\text{independent trials under same condition}}$$

In the probabilistic approach we are following

- Rule A is recovered immediately (under the assumption of equiprobability, when it applies).
- Rule B results from a theorem of Probability Theory (under well defined assumptions):
  ⇒ Laplace’s rule of succession (see later)

Mathematics of beliefs

The good news:

> *The basic laws of degrees of belief are the same we get from the inventory of favorable and possible cases, or from events occurred in the past.*

It can be proved that

> *the requirement of coherence leads to the famous 4 basic rules* $\Longrightarrow$

[Details skipped. . .]

Basic rules of probability

1. $0 \le P(A \mid I) \le 1$
2. $P(\Omega \mid I) = 1$
3. $P(A \cup B \mid I) = P(A \mid I) + P(B \mid I) \quad [\text{if } P(A \cap B \mid I) = \emptyset]$
4. $P(A \cap B \mid I) = P(A \mid B, I) \cdot P(B \mid I) = P(B \mid A, I) \cdot P(A \mid I)$

Remember that probability is always conditional probability!

I is the background condition (related to information 'I'_s')

→ usually implicit (we only care about 're-conditioning')

Basic rules of probability

1. $0 \leq P(A \,|\, I) \leq 1$
2. $P(\Omega \,|\, I) = 1$
3. $P(A \cup B \,|\, I) = P(A \,|\, I) + P(B \,|\, I) \quad [\text{if } P(A \cap B \,|\, I) = \emptyset]$

4. $P(A \cap B \,|\, I) = P(A \,|\, B, I) \cdot P(B \,|\, I) = P(B \,|\, A, I) \cdot P(A \,|\, I)$

Remember that probability is always conditional probability!

I is the background condition (related to information 'I_s')

→ usually implicit (we only care about 're-conditioning')

Note: 4. does not define conditional probability.
(Probability is always conditional probability!)

Mathematics of beliefs

An even <u>better news</u>:

The fourth basic rule
can be fully exploited!

Mathematics of beliefs

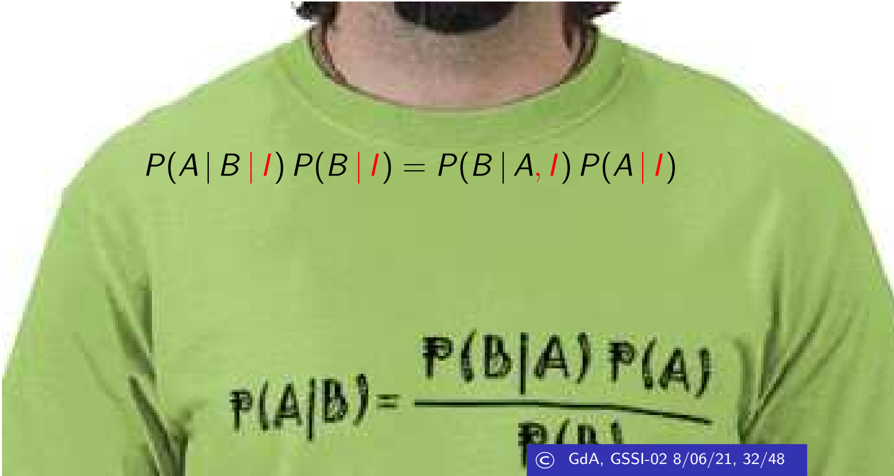
An even better news:

The fourth basic rule
can be fully exploited!

(Liberated by a curious ideology that forbids its use)

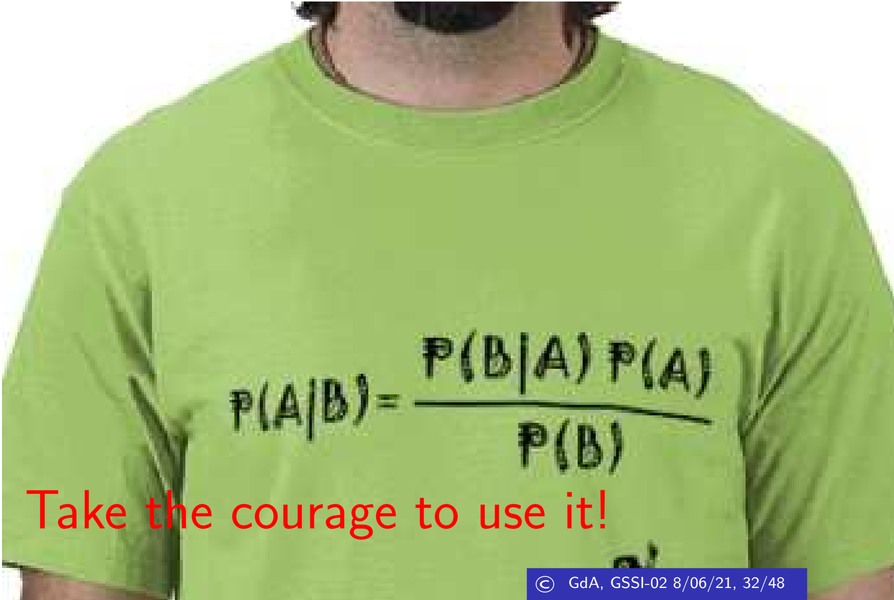
A simple, powerful formula

A simple, powerful formula



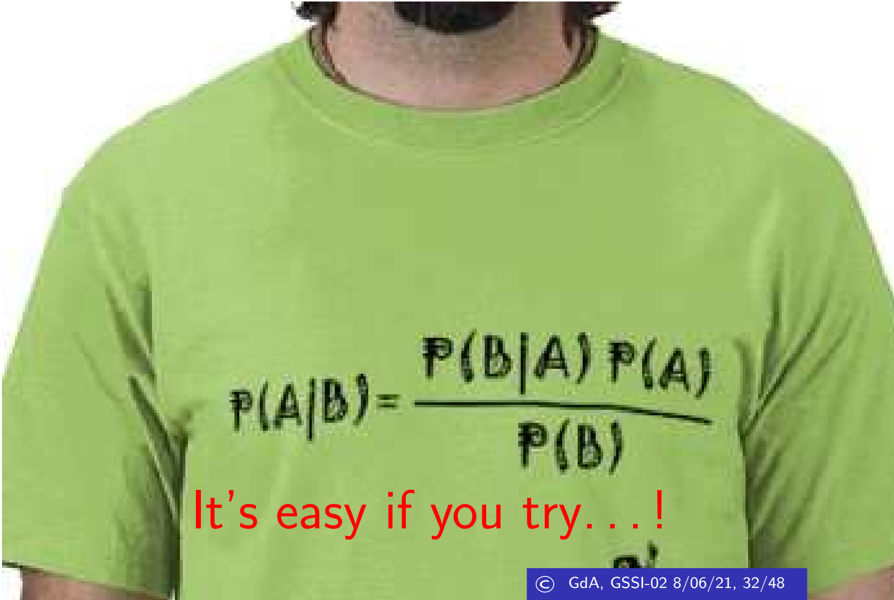
$$P(A \,|\, B \,|\, I)\, P(B \,|\, I) = P(B \,|\, A, I)\, P(A \,|\, I)$$

A simple, powerful formula



Take the courage to use it!

A simple, powerful formula



It's easy if you try. . . !

A simple, powerful formula

[Bayes Theorem]

A nice and powerful formula!

GdA and Allen Caldwell, Stellenbosch, South Africa, November 2013

[T-shirts kindly provided by Pangea Formazione]

Laplace's "Bayes Theorem"

"The greater the probability of an observed event given any one of a number of causes to which that event may be attributed, the greater the likelihood of that cause {given that event}.

$$P(C_i \,|\, E) \propto P(E \,|\, C_i)$$

Laplace's "Bayes Theorem"

"The greater the probability of an observed event given any one of a number of causes to which that event may be attributed, the greater the likelihood of that cause {given that event}. The probability of the existence of any one of these causes {given the event} is thus a fraction whose numerator is the probability of the event given the cause, and whose denominator is the sum of similar probabilities, summed over all causes.

$$P(C_i \mid E) = \frac{P(E \mid C_i)}{\sum_j P(E \mid C_j)}$$

Laplace's "Bayes Theorem"

"The greater the probability of an observed event given any one of a number of causes to which that event may be attributed, the greater the likelihood of that cause {given that event}. The probability of the existence of any one of these causes {given the event} is thus a fraction whose numerator is the probability of the event given the cause, and whose denominator is the sum of similar probabilities, summed over all causes. If the various causes are not equally probable *a priory*, it is necessary, instead of the probability of the event given each cause, to use the product of this probability and the *possibility of the cause itself*."

$$P(C_i \,|\, E) = \frac{P(E \,|\, C_i)\, P(C_i)}{\sum_j P(E \,|\, C_j)\, P(C_j)}$$

Laplace's "Bayes Theorem"

"The greater the probability of an observed event given any one of a number of causes to which that event may be attributed, the greater the likelihood of that cause {given that event}. The probability of the existence of any one of these causes {given the event} is thus a fraction whose numerator is the probability of the event given the cause, and whose denominator is the sum of similar probabilities, summed over all causes. If the various causes are not equally probable *a priory*, it is necessary, instead of the probability of the event given each cause, to use the product of this probability and the *possibility of the cause itself*."

$$P(C_i \,|\, E) = \frac{P(E \,|\, C_i)\, P(C_i)}{P(E)}$$

(*Philosophical Essai on Probabilities*)

[In general $P(E) = \sum_j P(E \,|\, C_j)\, P(C_j)$ (*weighted average*, with weigths being the probabilities of the conditions) if C_j form a *complete class of hypotheses*]

Laplace's "Bayes Theorem"

$$P(C_i \,|\, E) \;=\; \frac{P(E \,|\, C_i)\, P(C_i)}{P(E)} = \frac{P(E \,|\, C_i)\, P(C_i)}{\sum_j P(E \,|\, C_j)\, P(C_j)}$$

"This is the fundamental principle (*) of that branch of the analysis of chance that consists of reasoning a posteriori from events to causes"

(*) In his "Philosophical essay" Laplace calls 'principles' the 'fundamental rules'.

Laplace's "Bayes Theorem"

$$P(C_i \,|\, E) \;=\; \frac{P(E \,|\, C_i)\, P(C_i)}{P(E)} = \frac{P(E \,|\, C_i)\, P(C_i)}{\sum_j P(E \,|\, C_j)\, P(C_j)}$$

"This is the fundamental principle $^{(*)}$ of that branch of the analysis of chance that consists of reasoning a posteriori from events to causes"

(*) In his "Philosophical essay" Laplace calls 'principles' the 'fundamental rules'.

Note: denominator is just a normalization factor.

$$\Rightarrow \qquad P(C_i \,|\, E) \;\propto\; P(E \,|\, C_i)\, P(C_i)$$

Laplace's "Bayes Theorem"

$$P(C_i \mid E) \;=\; \frac{P(E \mid C_i)\, P(C_i)}{P(E)} = \frac{P(E \mid C_i)\, P(C_i)}{\sum_j P(E \mid C_j)\, P(C_j)}$$

"This is the fundamental principle $^{(*)}$ **of that branch of the analysis of chance that consists of reasoning a posteriori from events to causes"**

(*) In his "Philosophical essay" Laplace calls 'principles' the 'fundamental rules'.

Note: denominator is just a normalization factor.

$$\Rightarrow \qquad P(C_i \mid E) \;\propto\; P(E \mid C_i)\, P(C_i)$$

Most convenient way to remember Bayes theorem

Laplace's teaching

$$\frac{P(H_0 \,|\, \text{data})}{P(H_1 \,|\, \text{data})} \;=\; \frac{P(\text{data} \,|\, H_0)}{P(\text{data} \,|\, H_1)} \times \frac{P(H_0)}{P(H_1)}$$

- We should possibly use the data, rather then the test variables 'θ' (χ^2 etc);

  [although in some case 'sufficient summaries' do exist]

Laplace's teaching

$$\frac{P(H_0\,|\,\text{data})}{P(H_1\,|\,\text{data})} = \frac{P(\text{data}\,|\,H_0)}{P(\text{data}\,|\,H_1)} \times \frac{P(H_0)}{P(H_1)}$$

- We should possibly use the data, rather then the test variables 'θ' (χ^2 etc);
  [although in some case 'sufficient summaries' do exist]
- At least two hypotheses are needed!

Laplace's teaching

$$\frac{P(H_0 \,|\, \text{data})}{P(H_1 \,|\, \text{data})} = \frac{P(\text{data} \,|\, H_0)}{P(\text{data} \,|\, H_1)} \times \frac{P(H_0)}{P(H_1)}$$

- We should possibly use the data, rather then the test variables 'θ' (χ^2 etc);
  [although in some case 'sufficient summaries' do exist]
- At least two hypotheses are needed!
- ...and also how they appear belivable *a priori*!

Laplace's teaching

$$\frac{P(H_0 \,|\, \text{data})}{P(H_1 \,|\, \text{data})} = \frac{P(\text{data} \,|\, H_0)}{P(\text{data} \,|\, H_1)} \times \frac{P(H_0)}{P(H_1)}$$

- We should possibly use the data, rather then the test variables 'θ' (χ^2 etc);
  [although in some case 'sufficient summaries' do exist]
- At least two hypotheses are needed!
- ...and also how they appear belivable *a priori*!
- If $P(\text{data} \,|\, H_i) = 0$, it follows $P(H_i \,|\, \text{data}) = 0$:
  $\Rightarrow$ falsification (the 'serious' one) is a corollary of the theorem, rather than a principle.

Laplace's teaching

$$\frac{P(H_0 \,|\, \text{data})}{P(H_1 \,|\, \text{data})} = \frac{P(\text{data} \,|\, H_0)}{P(\text{data} \,|\, H_1)} \times \frac{P(H_0)}{P(H_1)}$$

- We should possibly use the data, rather then the test variables 'θ' (χ^2 etc);
  [although in some case 'sufficient summaries' do exist]
- At least two hypotheses are needed!
- ...and also how they appear belivable *a priori*!
- If $P(\text{data} \,|\, H_i) = 0$, it follows $P(H_i \,|\, \text{data}) = 0$:
  $\Rightarrow$ falsification (the 'serious' one) is a corollary of the theorem, rather than a principle.
- There is no conceptual problem with the fact that $P(\text{data} \,|\, H_1) \rightarrow 0$ (e.g. 10^{-37}), provided the ratio $P(\text{data} \,|\, H_0)/P(\text{data} \,|\, H_1)$ is not undefined.

Bayes factor ('likelihood ratio')

$$\frac{P(H_0\,|\,\text{data})}{P(H_1\,|\,\text{data})} \quad = \quad \frac{P(\text{data}\,|\,H_0)}{P(\text{data}\,|\,H_1)} \times \frac{P(H_0)}{P(H_1)}$$

Bayes factor ('likelihood ratio')

$$\frac{P(H_0\,|\,\text{data})}{P(H_1\,|\,\text{data})} \;=\; \frac{P(\text{data}\,|\,H_0)}{P(\text{data}\,|\,H_1)} \times \frac{P(H_0)}{P(H_1)}$$

$$\text{Prob. ratio}\big|_{posterior} \;=\; \text{Bayes factor} \times \text{Prob. ratio}\big|_{prior}$$

(*prior*/*posterior* w.r.t. data)

Bayes factor ('likelihood ratio')

$$\frac{P(H_0\,|\,\text{data})}{P(H_1\,|\,\text{data})} \;=\; \frac{P(\text{data}\,|\,H_0)}{P(\text{data}\,|\,H_1)} \times \frac{P(H_0)}{P(H_1)}$$

$$\text{Prob. ratio}\big|_{posterior} \;=\; \text{Bayes factor} \times \text{Prob. ratio}\big|_{prior}$$

(*prior*/*posterior* w.r.t. data)

If H_0 and H_1 are 'complementary', that is $H_1 = \overline{H}_0$, then

posterior odds = Bayes factor × prior odds

Application to the Aids test problem

(Left as exercise)

Apply this reasoning to the Aids test problem (Italian citizen chosen at random!) *taking* a number of HIV infected Italians of ≈ 100k:

Application to the Aids test problem

(Left as exercise)

Apply this reasoning to the Aids test problem (Italian citizen chosen at random!) *taking* a number of HIV infected Italians of ≈ 100k:

1. use the 'standard' Bayes theorem formula;

Application to the Aids test problem

(Left as exercise)

Apply this reasoning to the Aids test problem (Italian citizen chosen at random!) *taking* a number of HIV infected Italians of $\approx$ 100k:

1. use the 'standard' Bayes theorem formula;
2. use the Bayes factor;

Application to the Aids test problem

(Left as exercise)

Apply this reasoning to the Aids test problem (Italian citizen chosen at random!) *taking* a number of HIV infected Italians of ≈ 100k:

1. use the 'standard' Bayes theorem formula;
2. use the Bayes factor;
3. try to vary the assumed number of infected Italians
   - by $\pm 10\,\%$;
   - by $\pm 30\,\%$;
   - by $\pm 50\,\%$.

Application to the Aids test problem

(Left as exercise)

Apply this reasoning to the Aids test problem (Italian citizen chosen at random!) *taking* a number of HIV infected Italians of $\approx$ 100k:

1. use the 'standard' Bayes theorem formula;
2. use the Bayes factor;
3. try to vary the assumed number of infected Italians
   - by $\pm 10\,\%$;
   - by $\pm 30\,\%$;
   - by $\pm 50\,\%$.

And, needless to say, try to think to Covid-19 test issues:

Application to the Aids test problem

(Left as exercise)

Apply this reasoning to the Aids test problem (Italian citizen chosen at random!) *taking* a number of HIV infected Italians of ≈ 100k:

1. use the 'standard' Bayes theorem formula;
2. use the Bayes factor;
3. try to vary the assumed number of infected Italians
   - by $\pm 10\,\%$;
   - by $\pm 30\,\%$;
   - by $\pm 50\,\%$.

And, needless to say, try to think to Covid-19 test issues:

- dependence on priors;

Application to the Aids test problem

(Left as exercise)

Apply this reasoning to the Aids test problem (Italian citizen chosen at random!) *taking* a number of HIV infected Italians of ≈ 100k:

1. use the 'standard' Bayes theorem formula;
2. use the Bayes factor;
3. try to vary the assumed number of infected Italians
   - by $\pm 10\,\%$;
   - by $\pm 30\,\%$;
   - by $\pm 50\,\%$.

And, needless to say, try to think to Covid-19 test issues:

- dependence on priors;
- dependence on the fact that the test performances unavoidably some degree of uncertanty.

But statistical tests do work!

Someone would object that p-values and, in general, ‘hypothesis tests’ *usually* do work!

But statistical tests do work!

Someone would object that p-values and, in general, ‘hypothesis tests’ *usually* do work!

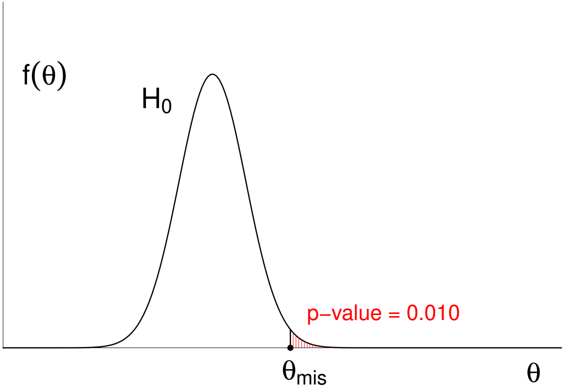
- Certainly! I agree!
  As it *usually* work overtakes in curve
  on remote mountain road!

But statistical tests do work!

Someone would object that p-values and, in general, 'hypothesis tests' *usually* do work!

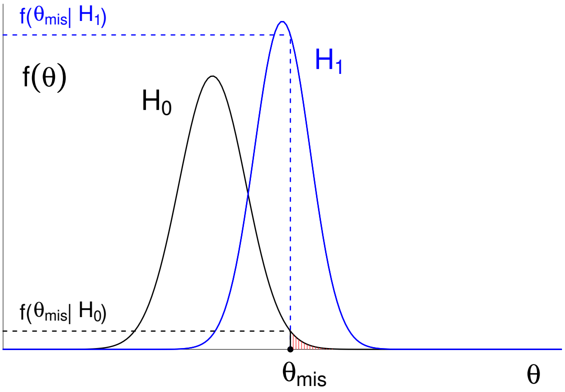
- Certainly! I agree!
  As it *usually* work overtakes in curve
  on remote mountain road!
- But now we are also able to explain the reason.

But statistical tests do work!



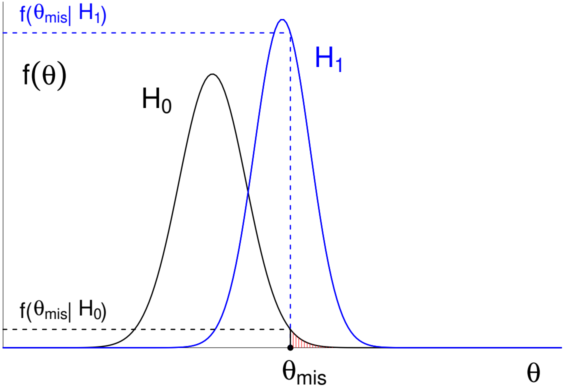
Why should the observation of θ_{mis} should diminish our confidence on H_0?

But statistical tests do work!



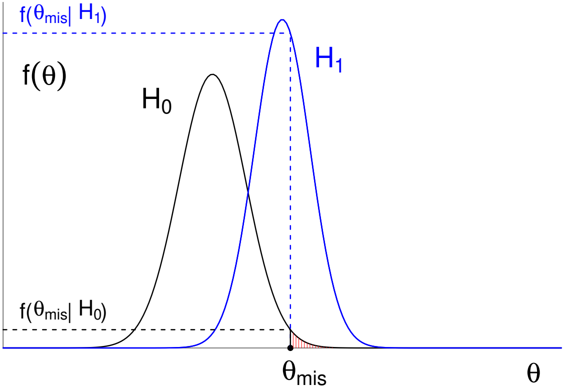
Because *often* we give *some chance* to a possible alternative hypothesis H_1, even if we are not able to exactly formulate it.

But statistical tests do work!



Indeed, what really matters <u>is not</u> the area to the right of θ_{mis}.
What matters is the ratio of $f(\theta_{mis}\,|\,H_1)$ to $f(\theta_{mis}\,|\,H_0)$!
$\Rightarrow$ to a 'small' area it corresponds a 'small' $f(\theta_{mis}\,|\,H_0)$.

But statistical tests do work!



But is the alternative hypothesis H_1 is unconceivable, or hardly believable, the 'smallness' of the area is irrelevant

Telling it with Gauss' words

A quote from the Princeps Mathematicorum
(Prince of Mathematicians) is a must.

Telling it with Gauss' words

A quote from the Princeps Mathematicorum
(Prince of Mathematicians) is a must.

$$P(C_i \mid \text{data}) = \frac{P(\text{data} \mid C_i)}{P(\text{data})} \, P_0(C_i)$$

Telling it with Gauss' words

A quote from the Princeps Mathematicorum (Prince of Mathematicians) is a must.

$$P(C_i \,|\, \text{data}) = \frac{P(\text{data} \,|\, C_i)}{P(\text{data})}\, P_0(C_i)$$

"post illa observationes" *"ante illa observationes"*

(Gauss)

Telling it with Gauss' words

A quote from the Princeps Mathematicorum
(Prince of Mathematicians) is a must.

$$P(C_i\,|\,\text{data}) = \frac{P(\text{data}\,|\,C_i)}{P(\text{data})}\,P_0(C_i)$$

"post illa observationes" *"ante illa observationes"*

(Gauss)

Arguments used to derive Gaussian distribution

- $f(\mu\,|\,\{x\}) \propto f(\{x\}\,|\,\mu)\cdot f_0(\mu)$
- $f_0(\mu)$ 'flat' (all values a priory equally possible)
- posterior maximized at $\mu = \overline{x}$

The Gauss' Bayes Factor

It might be curious to learn that Gauss had proved, **with emphasis**, the rule to update the ratio of probabilities of complementary hypotheses, in the light of an observed event which could be due to either of them.

The Gauss' Bayes Factor

It might be curious to learn that Gauss had proved, **with emphasis**, the rule to update the ratio of probabilities of complementary hypotheses, in the light of an observed event which could be due to either of them.

Although he focused on *a priori* equally probable hypotheses (explicitely stated!), in order to solve the problem on which he was interested in, the theorem can be easily extended to the general case.

The Gauss' Bayes Factor

It might be curious to learn that Gauss had proved, **with emphasis**, the rule to update the ratio of probabilities of complementary hypotheses, in the light of an observed event which could be due to either of them.

Although he focused on *a priori* equally probable hypotheses (explicitely stated!), in order to solve the problem on which he was interested in, the theorem can be easily extended to the general case.

And the resulting factor turns out to be what is presently known as Bayes Factor.

The Gauss' Bayes Factor

It might be curious to learn that Gauss had proved, **with emphasis**, the rule to update the ratio of probabilities of complementary hypotheses, in the light of an observed event which could be due to either of them.

Although he focused on *a priori* equally probable hypotheses (explicitely stated!), in order to solve the problem on which he was interested in, the theorem can be easily extended to the general case.

And the resulting factor turns out to be what is presently known as Bayes Factor.

⇒ `arXiv:2003.10878 [math.HO]`

The Gauss' Bayes Factor

It might be curious to learn that Gauss had proved, **with emphasis**, the rule to update the ratio of probabilities of complementary hypotheses, in the light of an observed event which could be due to either of them.

Although he focused on *a priori* equally probable hypotheses (explicitely stated!), in order to solve the problem on which he was interested in, the theorem can be easily extended to the general case.

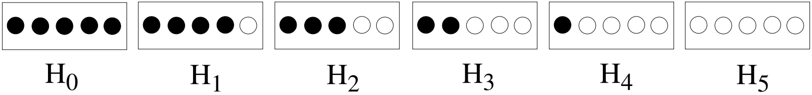
And the resulting factor turns out to be what is presently known as Bayes Factor.

$\Rightarrow$ `arXiv:2003.10878 [math.HO]`

(And, by the way, Enrico Fermi derived analysis tools based on *his* Bayes Theorem...

$\Rightarrow$ `arXiv:physics/0509080 [physics.hist-ph]`)

Application to the six box problem



Remind:

- $E_1 =$ White
- $E_2 =$ Black

Collecting the pieces of information we need

Our tool:

$$P(H_j \mid E_i, I) = \frac{P(E_i \mid H_j, I)}{P(E_i \mid I)} P(H_j \mid I)$$

Collecting the pieces of information we need

Our tool:

$$P(H_j \mid E_i, I) = \frac{P(E_i \mid H_j, I)}{P(E_i \mid I)} P(H_j \mid I)$$

- $P(H_j \mid I) = 1/6$

Collecting the pieces of information we need

Our tool:

$$P(H_j \mid E_i, I) = \frac{P(E_i \mid H_j, I)}{P(E_i \mid I)} P(H_j \mid I)$$

- $P(H_j \mid I) = 1/6$
- $P(E_i \mid I) = 1/2$

Collecting the pieces of information we need

Our tool:

$$\boxed{P(H_j \,|\, E_i, I) = \frac{P(E_i \,|\, H_j, I)}{P(E_i \,|\, I)}\, P(H_j \,|\, I)}$$

- $P(H_j \,|\, I) = 1/6$
- $P(E_i \,|\, I) = 1/2$
- $P(E_i \,|\, H_j,\, I)$:

$$\begin{aligned} P(E_1 \,|\, H_j,\, I) &= j/5 \\ P(E_2 \,|\, H_j,\, I) &= (5-j)/5 \end{aligned}$$

Collecting the pieces of information we need

Our tool:

$$P(H_j \mid E_i, I) = \frac{P(E_i \mid H_j, I)}{P(E_i \mid I)}\, P(H_j \mid I)$$

- $P(H_j \mid I) = 1/6$
- $P(E_i \mid I) = 1/2$
- $P(E_i \mid H_j, I)$:

$$\begin{aligned} P(E_1 \mid H_j, I) &= j/5 \\ P(E_2 \mid H_j, I) &= (5-j)/5 \end{aligned}$$

Our prior belief about H_j

Collecting the pieces of information we need

Our tool:

$$P(H_j \,|\, E_i, I) = \frac{P(E_i \,|\, H_j, I)}{P(E_i \,|\, I)}\, P(H_j \,|\, I)$$

- $P(H_j \,|\, I) = 1/6$
- $P(E_i \,|\, I) = 1/2$
- $P(E_i \,|\, H_j,\, I)$:

$$\begin{aligned} P(E_1 \,|\, H_j,\, I) &= j/5 \\ P(E_2 \,|\, H_j,\, I) &= (5-j)/5 \end{aligned}$$

Probability of E_i under a well defined hypothesis H_j
It corresponds to the 'response of the apparatus' in measurements.
→ likelihood (traditional, rather confusing name!)

Collecting the pieces of information we need

Our tool:

$$P(H_j \,|\, E_i, I) = \frac{P(E_i \,|\, H_j, I)}{P(E_i \,|\, I)} \, P(H_j \,|\, I)$$

- $P(H_j \,|\, I) = 1/6$
- $P(E_i \,|\, I) = 1/2$
- $P(E_i \,|\, H_j, I)$:

$$\begin{aligned} P(E_1 \,|\, H_j, I) &= j/5 \\ P(E_2 \,|\, H_j, I) &= (5-j)/5 \end{aligned}$$

Probability of E_i taking account all possible H_j

$\rightarrow$ How much we are confident that E_i will occur.

Collecting the pieces of information we need

Our tool:

$$P(H_j \,|\, E_i, I) = \frac{P(E_i \,|\, H_j, I)}{P(E_i \,|\, I)}\, P(H_j \,|\, I)$$

- $P(H_j \,|\, I) = 1/6$
- $P(E_i \,|\, I) = 1/2$
- $P(E_i \,|\, H_j, I)$:

$$\begin{aligned} P(E_1 \,|\, H_j, I) &= j/5 \\ P(E_2 \,|\, H_j, I) &= (5-j)/5 \end{aligned}$$

Probability of E_i taking account all possible H_j
→ How much we are confident that E_i will occur.
(taking into account all possible hypotheses H_j)

Collecting the pieces of information we need

Our tool:

$$P(H_j \,|\, E_i, I) = \frac{P(E_i \,|\, H_j, I)}{P(E_i \,|\, I)} \, P(H_j \,|\, I)$$

- $P(H_j \,|\, I) = 1/6$
- $P(E_i \,|\, I) = 1/2$
- $P(E_i \,|\, H_j, I)$:

$$\begin{aligned} P(E_1 \,|\, H_j, I) &= j/5 \\ P(E_2 \,|\, H_j, I) &= (5-j)/5 \end{aligned}$$

But it easy to prove that $P(E_i \,|\, I)$ is related to the other ingredients, usually easier to 'measure' or to assess somehow, though vaguely

Collecting the pieces of information we need

Our tool:

$$P(H_j \mid E_i, I) = \frac{P(E_i \mid H_j, I)}{P(E_i \mid I)} P(H_j \mid I)$$

- $P(H_j \mid I) = 1/6$
- $P(E_i \mid I) = 1/2$
- $P(E_i \mid H_j, I)$:

$$\begin{aligned} P(E_1 \mid H_j, I) &= j/5 \\ P(E_2 \mid H_j, I) &= (5-j)/5 \end{aligned}$$

But it easy to prove that $P(E_i \mid I)$ is related to the other ingredients, usually easier to 'measure' or to assess somehow, though vaguely

'decomposition law': $P(E_i \mid I) = \sum_j P(E_i \mid H_j, I) \cdot P(H_j \mid I)$
($\rightarrow$ Easy to check that it gives $P(E_i \mid I) = 1/2$ in our case).

Collecting the pieces of information we need

Our tool:

$$P(H_j \mid E_i, I) = \frac{P(E_i \mid H_j, I) \cdot P(H_j \mid I)}{\sum_j P(E_i \mid H_j, I) \cdot P(H_j \mid I)}$$

- $P(H_j \mid I) = 1/6$
- $P(E_i \mid I) = \sum_j P(E_i \mid H_j, I) \cdot P(H_j \mid I)$
- $P(E_i \mid H_j, I)$:

$$\begin{aligned} P(E_1 \mid H_j, I) &= j/5 \\ P(E_2 \mid H_j, I) &= (5-j)/5 \end{aligned}$$

We are ready!

$\longrightarrow$ Let's play with our toy

We are ready

Now that we have set up our formalism, let's play a little

- analyse real data
- some simulations
- make variations

We are ready

Now that we have set up our formalism, let's play a little

- analyse real data
- some simulations
- make variations

Let's play!

- Hugin Expert (Lite – demo version);
- R scripts

Playing with the six boxes

Learning by simulations

- History of $P(H_j \,|\, \text{obs. sequence})$.

Playing with the six boxes

Learning by simulations

- History of $P(H_j \,|\, \text{obs. sequence})$.
- History of $P(B/W \,|\, \text{obs. sequence})$.

Playing with the six boxes

Learning by simulations

- History of $P(H_j \,|\, \text{obs. sequence})$.
- History of $P(B/W \,|\, \text{obs. sequence})$.
- Comparison of the $P(B/W \,|\, \text{obs. sequence})$ with the relative frequency with the color has occurred in the past ("probability evaluated by relative frequency").

Playing with the six boxes

Learning by simulations

- History of $P(H_j \,|\, \text{obs. sequence})$.
- History of $P(B/W \,|\, \text{obs. sequence})$.
- Comparison of the $P(B/W \,|\, \text{obs. sequence})$ with the relative frequency with the color has occurred in the past ("probability evaluated by relative frequency").
  - Why does the Bayesian solution performs better?

Playing with the six boxes

Learning by simulations

- History of $P(H_j \,|\, \text{obs. sequence})$.
- History of $P(B/W \,|\, \text{obs. sequence})$.
- Comparison of the $P(B/W \,|\, \text{obs. sequence})$ with the relative frequency with the color has occurred in the past ("probability evaluated by relative frequency").
  - Why does the Bayesian solution performs better?

  → It takes into account at the best all available information. (The frequency based answer is, at most, the solution to a different problem...)

Playing with the six boxes

Learning by simulations

- History of $P(H_j \,|\, \text{obs. sequence})$.
- History of $P(B/W \,|\, \text{obs. sequence})$.
- Comparison of the $P(B/W \,|\, \text{obs. sequence})$ with the relative frequency with the color has occurred in the past ("probability evaluated by relative frequency").
  - Why does the Bayesian solution performs better?

  → It takes into account at the best all available information. (The frequency based answer is, at most, the solution to a different problem...)
- Comparison of $P(H_j \,|\, \text{obs. sequence})$ with frequentistic methods?

Playing with the six boxes

Learning by simulations

- History of $P(H_j \,|\, \text{obs. sequence})$.
- History of $P(B/W \,|\, \text{obs. sequence})$.
- Comparison of the $P(B/W \,|\, \text{obs. sequence})$ with the relative frequency with the color has occurred in the past ("probability evaluated by relative frequency").
  - Why does the Bayesian solution performs better?
  - → It takes into account at the best all available information. (The frequency based answer is, at most, the solution to a different problem...)
- Comparison of $P(H_j \,|\, \text{obs. sequence})$ with frequentistic methods?

NO!

Playing with the six boxes

Learning by simulations

- History of $P(H_j \,|\, \text{obs. sequence})$.
- History of $P(B/W \,|\, \text{obs. sequence})$.
- Comparison of the $P(B/W \,|\, \text{obs. sequence})$ with the relative frequency with the color has occurred in the past ("probability evaluated by relative frequency").
  - Why does the Bayesian solution performs better?
  - → It takes into account at the best all available information. (The frequency based answer is, at most, the solution to a different problem...)
- Comparison of $P(H_j \,|\, \text{obs. sequence})$ with frequentistic methods?

NO!

- Don't even think: frequentists refuse to assign probabilities to hypotheses (in general), to causes, to true values, etc.
  (And you have seen the results. . .)

How does it work?

Simple case (no reporter/composition/etc. complications)

How does it work?

Simple case (no reporter/composition/etc. complications)

- Update probabilities of hypotheses (cause, Box): *inference*:

$$P^{(n)}(B_j) \propto P(E_i^{(n)} \mid B_j) \cdot P^{(n-1)}(B_j)$$

How does it work?

Simple case (no reporter/composition/etc. complications)

- Update probabilities of hypotheses (cause, Box): *inference*:

$$P^{(n)}(B_j) \propto P(E_i^{(n)} \mid B_j) \cdot P^{(n-1)}(B_j)$$

- Update probabilities of next extraction: *prediction*:

$$P^{(n+1)}(E_i) = \sum_j P(E_i \mid B_j) \cdot P^{(n)}(B_j)$$

How does it work?

General case (more complicate 'network')

How does it work?

General case (more complicate 'network')
for example including uncertain Composition (C) and a Reporter for each estraction ($\underline{R}$):

How does it work?

General case (more complicate 'network')

for example including uncertain Composition (C) and a Reporter for each estraction ($\underline{R}$):

- Write down the joint distribution of all variables in the game:

$$P(C, B, \underline{E}, \underline{R})$$

How does it work?

General case (more complicate 'network')
for example including uncertain Composition (C) and a Reporter for each estraction ($\underline{R}$):

- Write down the joint distribution of all variables in the game:

$$P(C, B, \underline{E}, \underline{R})$$

$$\begin{aligned} \underline{E}&: \; E_i^{(1)}, E_i^{(2)}, E_i^{(3)}, \ldots \\ \underline{R}&: \; R_i^{(1)}, R_i^{(2)}, R_i^{(3)}, \ldots \end{aligned}$$

How does it work?

General case (more complicate 'network')
for example including uncertain Composition (C) and a Reporter for each estraction ($\underline{R}$):

- Write down the joint distribution of all variables in the game:

$$P(C, B, \underline{E}, \underline{R})$$

$$\begin{array}{ll} \underline{E}: & E_i^{(1)}, E_i^{(2)}, E_i^{(3)}, \ldots \\ \underline{R}: & R_i^{(1)}, R_i^{(2)}, R_i^{(3)}, \ldots \end{array}$$

- Condition on the 'observations':

$$P(C, B, \underline{E}, \underline{R}^{(k>n)} \,|\, \underline{R}^{(k\leq n)}) = \frac{P(C, B, \underline{E}, \underline{R})}{P(\underline{R}^{(k\leq n)})}$$

How does it work?

General case (more complicate 'network')
for example including uncertain Composition (C) and a Reporter for each estraction ($\underline{R}$):

- Write down the joint distribution of all variables in the game:

$$P(C, B, \underline{E}, \underline{R})$$

$$\begin{array}{ll} \underline{E}: & E_i^{(1)}, E_i^{(2)}, E_i^{(3)}, \ldots \\ \underline{R}: & R_i^{(1)}, R_i^{(2)}, R_i^{(3)}, \ldots \end{array}$$

- Condition on the 'observations':

$$P(C, B, \underline{E}, \underline{R}^{(k>n)} \,|\, \underline{R}^{(k\leq n)}) = \frac{P(C, B, \underline{E}, \underline{R})}{P(\underline{R}^{(k\leq n)})}$$

No real distinction between inference and prediction

How does it work?

General case (more complicate 'network')
for example including uncertain Composition (C) and a Reporter for each estraction ($\underline{R}$):

- Write down the joint distribution of all variables in the game:

$$P(C, B, \underline{E}, \underline{R})$$

$$\begin{aligned} \underline{E}&: \; E_i^{(1)}, E_i^{(2)}, E_i^{(3)}, \ldots \\ \underline{R}&: \; R_i^{(1)}, R_i^{(2)}, R_i^{(3)}, \ldots \end{aligned}$$

- Condition on the 'observations':

$$P(C, B, \underline{E}, \underline{R}^{(k>n)} \,|\, \underline{R}^{(k\leq n)}) = \frac{P(C, B, \underline{E}, \underline{R})}{P(\underline{R}^{(k\leq n)})}$$

No real distinction between inference and prediction
(We shall see it later in the case of *continuous distributions*)

The End